

UNCLASSIFIED

AD 4 3 6 8 6 8

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

64-12

436868

11 Rept. no. 1178

1

9

MONTHLY TECHNICAL REPORT.

For Dec 63.

6

FORT DETRICK THESAURUS PROJECT.

11

8 Jan 1964

15

Contract DA 18-064-AMC-132(A)

436868

Prepared for:

U. S. Army Biological Laboratories
Fort Detrick, Maryland

Prepared by:

General Electric Company
Information Systems Operation
4901 Fairmont Avenue
Washington, D. C. 20044

DDC
APR 28 1964
TISA

HU NO. _____
DDC FILE COPY

354650
6

SP #1.10

FORT DETRICK THESAURUS PROJECT

INTRODUCTION

This informal report presents a summary of the technical progress for the month of December 1963.

PROGRESS IN DECEMBER

Work during December fell into the following areas:

- o Planning of Phase II indexing effort.
- o Planning of Phase II thesaurus effort.
- o Indexing of Fort Detrick documents.
- o Modification of display of Frequently Used Descriptors.

Planning of Phase II indexing effort

On 4 December 1963 at Fort Detrick, the following conclusions were reached at a meeting between Fort Detrick and General Electric personnel:

Recorded information - For each document indexed, the recorded information will include (1) document accession number, (2) document security classification, (3) code number of each descriptor selected for indexing, (4) total number of descriptors selected for indexing the document, and (5) the top-level Display I descriptor, such as Immunity, Pathology, etc., associated with each descriptor selected for indexing.

Document accession number - Each document will be uniquely identified by its Fort Detrick five-digit accession number. Fort Detrick personnel will make a decision in regard to the use of a sixth character for designating the security classification of a document.

Descriptor code - Each unique thesaurus descriptor will be assigned a unique five-digit code. Thus, the same code will be used to represent a given descriptor regardless of the number of times that it appears within the thesaurus. In many instances combined descriptors, such as "Plasma membrane", will appear in the thesaurus and will be able to be selected for indexing. In such an instance the code for "Plasma" and the code for "membrane" will be used for indexing. If a particular combination of terms is frequently used, consideration will be given to assigning a unique five-digit code to the combination. This code and the codes for the individual descriptors will be used for indexing purposes.

Machine readable medium for presenting document number-descriptor records - A Friden 8-channel punched paper tape will be used for presenting the document number-descriptor records unless Fort Detrick and General Electric mutually agree to modify this plan.

Format of document number-descriptor records - The format of the index information will be a series of records each consisting of a document accession number followed by the code numbers of the descriptors having been selected for indexing the document. Records of classified and unclassified documents will appear on the same tape. Spaces required between records and within records and any required special characters will be specified by Fort Detrick.

Statistical data on document-descriptor relationships - The types of data desired will be considered by Fort Detrick personnel. Plans are to hold a Fort Detrick-General Electric meeting for mutually agreeing upon the types of data to be presented.

Indexing approach - The initial step will be to index about 100 representative Fort Detrick documents and to discuss the results with Fort Detrick personnel. The experience gained during this indexing period will be a guide for considering modifications to the Display II descriptors. An inverted index file for manual operation will be prepared in order to conduct tests on the search capability resulting from the indexing. Need to maintain the inverted file after the 100 documents have been indexed will be discussed.

Planning of Phase II thesaurus effort

At the 4 December 1963 meeting at Fort Detrick, the following conclusions were reached:

Thesaurus format - As discussed throughout the project, there will be three thesaurus displays. One will be the Total Grouping of Descriptors submitted in Phase I. Modification to the present display will be based upon (1) information gained during indexing and (2) guidance provided by Fort Detrick personnel. Based upon these same factors, a more radical modification will be made to the display of Frequently Used Descriptors which was also prepared during Phase I. Included in this display section will be definitions of the descriptors tabulated in the same order as the displayed descriptors, tabulated alphabetically, or tabulated by some other format to be mutually agreed upon. A third display, Alphabetic Display of Descriptors, will include "Use", "Also see", etc. designations. Consideration is being given to flag a first or third display descriptor if it also appears in the display of Frequently Used Descriptors.

Machine readable medium for presenting thesaurus information - A Friden 8-channel punched paper tape will be used for presenting the required thesaurus information unless Fort Detrick and General Electric mutually agree to modify this plan.

Required machine readable thesaurus information - The required thesaurus

information is an alphabetic listing of the descriptors included in the thesaurus together with the corresponding descriptor codes. Details of the tape format will be specified by Fort Detrick.

Descriptor record - Definitions of the descriptors appearing in the Frequently Used Descriptors section will be considered justification for including the descriptors in the thesaurus. Justification for the selection of other descriptors for the other displays will require a listing of the respective sources from which the descriptors were selected. It will not be necessary to include the number of the page from which a particular descriptor was selected.

Indexing of Fort Detrick documents

Eighteen of the 100 representative Fort Detrick documents have been indexed. The documents that were processed were selected at random from a box of documents furnished by Fort Detrick.

A form was designed for recording indexing data. Space is provided for recording the Fort Detrick accession number appearing on the document, the document's security classification, the indexer's initials, and the date on which the document was indexed. Space is also provided for recording selected indexing descriptors in natural language and by a five-digit code. During the actual indexing each selected descriptor was listed in natural language as it appears in the thesaurus. The corresponding five-digit code was not recorded since codes have not been assigned to the descriptors as yet. The appropriate code numbers will be inserted at a later date.

When combined terms, such as "mathematical modeling", are selected from the thesaurus for indexing, the descriptor "mathematical" and the descriptor "modeling" are recorded as independent descriptors and are bracketed to indicate the combinations which appears in the thesaurus. When a descriptor not appearing in the thesaurus is used for indexing, an asterisk is placed in front of the descriptor to so indicate. The form also provides space for recording the group and page number from which a descriptor has been selected in Display I and Display II.

Modification of display of Frequently Used Descriptors

Experience acquired in the indexing of the small sample of documents confirms the need to modify the display of Frequently Used Descriptors. Additions and deletions must be made to the descriptors tabulated under the various main headings. Also, some degree of structuring would be desirable. During the indexing, notes are being made to record ideas or definite suggestions for improving the displays.

Some effort has been devoted to experiment with modifications to the display of Frequently Used Descriptors. In the area of Pathology, for example, concepts associated with Disease were considered. A structure involving Disease is tabulated in APPENDIX A. The structure is not finalized but can be used to illustrate the direction in which improvement can be realized.

APPENDIX A

DESCRIPTOR STRUCTURE INVOLVING DISEASE

Disease

Causative agent See Agent
 Control (of)
 Decontamination
 Immunization See Immunity
 Protective devices
 Reservoirs
 Sanitation
 Vectors
 Diagnosis (of)
 Symptoms
 Tests
 Distribution (of) See Epidemiology
 Immunity (to) See Immunity
 Names (of)
 Human (diseases)

Plant (diseases)

Veterinary (diseases)

Disease (continued)

Pathogenesis
 Types (of)
 Human
 Plant
 Veterinary

 Incapacitating
 Lethal

 Circulatory
 Deficiency
 Infectious
 Local
 Malignant
 Nervous
 Respiratory
 Serum
 Specific
 Systemic
 Spread (of)
 Symptoms (of)
 Therapy (for)